Appendix A

NUSF-64

A Method for Developing Cost Caps for NUSF Funding

The NPSC is committed to supporting service in high cost areas through its universal service funding mechanism. To ensure that as many Nebraskans as possible are served, the NUSF funding model curtails payments when companies are making an overall return in excess of some predetermined amount. Part of the process for estimating a company's return involves reviewing reported costs. The NPSC is not interested in implementing an in-depth cost review for every company requesting universal service funds. However, the Commission does want to ensure that costs are reasonable. Toward this end, the NPSC proposes a methodology to estimate a company-specific threshold on allowable total costs. Companies that report costs below the threshold would face no further scrutiny from the Commission. Companies reporting costs above the threshold would need to be approved by the Commission.

The NPSC will use information previously reported by companies as the basis for forming its forecast cost threshold. Form M data for the years 2000 through 2005 are collected for all reporting companies. A sample of data is developed that contains total cost, the number of square miles served, the number of households and access lines, and the value of plant in service by company. Total cost is defined as the sum of reported costs in accounts 6110, 6120, 6210, 6220, 6230, 6310, 6410, 6510, 6530, 6540, 6610, 6620, 6710, and 6720.   The sample contains 141 observations and includes all reports where complete data were available.

Regression analysis is used to estimate total cost as a function of several regressors. The model is specified as:

$$\begin{aligned} TC_{i,t} &= \beta_0 + \beta_1 SqMi_{i,t} + \beta_2 Households_{i,t} + \beta_3 AccLines_{i,t} \\ &\quad + \beta_4 Households_{i,t} * AccLines_{i,t} + \beta_5 PlantInSvc_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{1.1}$$

where $TC_{i,t}$ is total cost as reported by company i in year t, $SqMi_{i,t}$ is square miles served, $Households_{i,t}$ and $AccLines_{i,t}$ are the number of households and access lines, respectively, $PlantInSvc_{i,t}$ is plant in service as reported by company i in year t, and $\varepsilon_{i,t}$ is an error term assumed iid. Regression results are used to forecast expected total cost as:

$$\begin{aligned} TC_{i,t} &= \widehat{\beta_0} + \widehat{\beta_1} SqMi_{i,t+1} + \widehat{\beta_2} Households_{i,t+1} + \widehat{\beta_3} AccLines_{i,t+1} \\ &\quad + \widehat{\beta_4} Households_{i,t+1} * AccLines_{i,t+1} + \widehat{\beta_5} PlantInSvc_{i,t+1} \end{aligned} . \tag{1.2}$$

If the vector $\underline{X}_{i,t+1}$ contains forecast regressors and the matrix $\mathbf{\Omega}$ is the covariance matrix for the parameters in equation (1.1), then s_f , the forecast standard deviation for any observation is:

$$s_f = \sqrt{s^2 + \underline{X}_{i,t+1}^{'} \Omega \underline{X}_{i,t+1}} \ , \tag{1.3}$$

where s^2, the variance of the regression, is defined as:

$$s^2 = \frac{1}{N-k}\sum(TC_{i,t} - \widehat{TC}_{i,t})^2\ , \qquad (1.4)$$

In equation (1.4), *N* is the number of observations used in the regression and *k* is the number of regressors including the constant. Using this information, the upper bound on non-reviewable costs can be calculated as:

$$TC_{i,t+1}^{Upper} = \widehat{TC}_{i,t+1} + t_\alpha s_f\ , \qquad (1.5)$$

where t_α is the parameter from a standard normal distribution that creates the one-sided confidence interval of (0.5 - α). We set the parameter α equal to 0.25 so that 75 percent of the distribution of forecast total cost lies below the upper threshold. This includes 50 percent of the distribution below the expected value plus another twenty five percent of the distribution above the expected value and below the upper boundary.

The estimated parameters and their standard errors are shown in Table 1, below. All of the coefficients are statistically significant at the 95 percent confidence level using a two-tailed test. The equation has an $\overline{R}^2$ = 0.99 and the f-statistic is significant at the 99 percent level.

Table 1
Total Cost as a Function of Regressors

Regressor	Coefficient	Standard Error
Constant	420,990	145. 4
SqMi	264.2	52.8
Households	252.3	75.2
AccLines	89.2	20.7
*Households*AccLines*	0.00015	7.15E-5
PlantInSvc	0.025	0.0091

The covariance matrix for the coefficients is shown in Table 2. The standard error of the regression is 1,460,806.

Table 2
Coefficient Covariance Matrix

	Constant	*SqMi*	*Households*	*AccLines*	*Households *AccLines*	*PlantInSvc*
Constant	2.11E+10	-2.93E+06	8.83E+05	-2.87E+05	-2.73E-01	-1.13E+02
SqMi	-2.93E+06	2.79E+03	-3.78E+02	-2.64E+01	8.28E-04	6.42E-03
Households	8.83E+05	-3.78E+02	5.65E+03	-1.46E+03	-4.94E-03	-6.66E-01
AccLines	-2.87E+05	-2.64E+01	-1.46E+03	4.27E+02	1.15E-03	1.69E-01
Households *AccLines	-2.73E-01	8.28E-04	-4.94E-03	1.15E-03	5.11E-09	5.52E-07
PlantInSvc	-1.13E+02	6.42E-03	-6.66E-01	1.69E-01	5.52E-07	8.26E-05

A number of statistics can be used to measure how closely this forecast method tracks the actual data. Theil's inequality, typically symbolized as *U*, is one such measure. It is defined as:

$$U = \frac{\sqrt{\frac{1}{N}\sum (TC^f_{i,t} - TC^a_{i,t})^2}}{\sqrt{\frac{1}{N}\sum (TC^f_{i,t})^2} + \sqrt{\frac{1}{N}\sum (TC^a_{i,t})^2}}, \tag{1.6}$$

where $TC^f_{i,t}$ is forecasted total cost and $TC^a_{i,t}$ is actual total cost. Theil's inequality should take a value between zero and one with values closer to zero indicating a better forecast. Using the actual data in our sample and the forecast method, it is possible to calculate Theil's inequality for our sample of data. In this sample, Theil's inequality is equal to 0.027.

Theil's inequality can be decomposed into three parts, U^M which is called the bias proportion, U^S which is called the variance proportion, and U^C which is called the covariance proportion. Letting $\overline{Y}^f$, $\overline{Y}^a$, σ_f and σ_a represent the means and standard deviations of the forecasted and actual series, respectively, and ρ be their correlation coefficient defined as:

$$\rho = \frac{\sum (Y^f_{i,t} - \overline{Y}^f)(Y^a_{i,t} - \overline{Y}^a)}{\sigma_f \sigma_a N} \tag{1.7}$$

The proportions can then be defined as:

$$U^{M} = \frac{(\overline{Y}^{f} - \overline{Y}^{a})^{2}}{(1/N)\sum(Y_{i,t}^{f} - Y_{i,t}^{a})^{2}}, \tag{1.8}$$

$$U^{S} = \frac{(\sigma_{f} - \sigma_{a})^{2}}{(1/N)\sum(Y_{i,t}^{f} - Y_{i,t}^{a})^{2}}, \tag{1.9}$$

and

$$U^{C} = \frac{2(1-\rho)\sigma_{f}\sigma_{a}}{(1/N)\sum(Y_{i,t}^{f} - Y_{i,t}^{a})^{2}}. \tag{1.10}$$

The bias proportion (U^M) is an indication of systematic error and should be close to zero. In our sample, it is almost zero (1.5E-7). The variance proportion (U^S) shows the ability of the forecast to replicate the degree of variability in actual total costs and should be near zero as well. In our sample, it equals 0.0009. The covariance proportion (U^C) measures unsystematic error and should be close to one, which it is for our sample.

The parameter results from Table 1 and the boundary described in equation 1.5 can be used to set a threshold for triggering a cost review. On a going forward basis, companies will report to the Commission their Form M costs. Account data will be aggregated to get company actual total cost. In addition, access line counts, square miles and households served and plant in service will be combined with the parameter estimates from Table 1 to derive forecasted total cost and equation 1.5 will be used to determine the upper boundary for total costs.

If actual total cost is below the upper boundary, no further review will be done. However, if actual cost falls above the boundary, the Commission will determine the cause of the disparity from expected costs and would need to approve that variance.